

Structural bioinformatics

JADOPPT: java based AutoDock preparing and processing tool

Carlos García-Pérez^{1,2}, Rafael Peláez^{3,4,5,*}, Roberto Therón² and José Luis López-Pérez³

¹Centro de Biotecnología Genómica, Instituto Politécnico Nacional, Reynosa, Mexico and ²Departamento de Informática y Automática, ³Química Farmacéutica, ⁴Instituto de Investigación Biomédica de Salamanca (IBSAL) and ⁵Centro de Investigación de Enfermedades Tropicales de la Universidad de Salamanca (CIETUS), Salamanca, Spain

Associate Editor: Anna Tramontano

*To whom correspondence should be addressed

Received on May 12, 2016; revised on October 19, 2016; editorial decision on October 22 2016; accepted on October 25, 2016

Abstract

Motivation: AutoDock is a very popular software package for docking and virtual screening. However, currently it is hard work to visualize more than one result from the virtual screening at a time. To overcome this limitation we have designed JADOPPT, a tool for automatically preparing and processing multiple ligand-protein docked poses obtained from AutoDock. It allows the simultaneous visual assessment and comparison of multiple poses through clustering methods. Moreover, it permits the representation of reference ligands with known binding modes, binding site residues, highly scoring regions for the ligand, and the calculated binding energy of the best ranked results.

Availability and Implementation: JADOPPT, [supplementary material](#) (Case Studies 1 and 2) and video tutorials are available at <http://visualanalytics.land/cgarcia/JADOPPT.html>

Contacts: carlosgarcia@usal.es or pelaez@usal.es

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

The discovery of new drugs and chemicals with biological activity through virtual screening is a well-established methodology. Among the extant docking software tools, AutoDock (Morris *et al.*, 1998) is one of the most popular. However, virtual screening campaigns produce a large number of docking poses that require costly visual efforts to determine the quality of the target-ligand interactions. In this context, several tools have been developed for the analysis of the AutoDock docking results, such as AutoDockTools (Morris *et al.*, 2009), BDT (Vaqué *et al.*, 2006), DOVIS (Zhang *et al.*, 2008) and plugin access through PyMOL software (Lill and Danielson, 2011; Seeliger and Groot, 2010). JADOPPT is an alternative tool that allows highly interactive visual analysis, and provides means for refinement of docking studies. It also overcomes the limitations of single molecule analysis by employing a clustering

methodological approach and allows comparisons of multiple dockings. Thus, JADOPPT represents a new application with diverse functionalities that is likely to be a valuable contribution in the field of drug discovery.

2 Methods

JADOPPT performs three main tasks. Firstly, it focuses on reducing the dimensionality problem by hierarchically selecting representative poses for each docked molecule. JADOPPT calculates the RMSD between the poses and automatically clusters them. The size of the clusters is determined by the RMSD threshold (default 2.0 Å). The pose with minimum energy within each cluster is selected as its representative. As a result, the dataset size is reduced while the information richness of the chemical sampling is preserved. Two kinds of

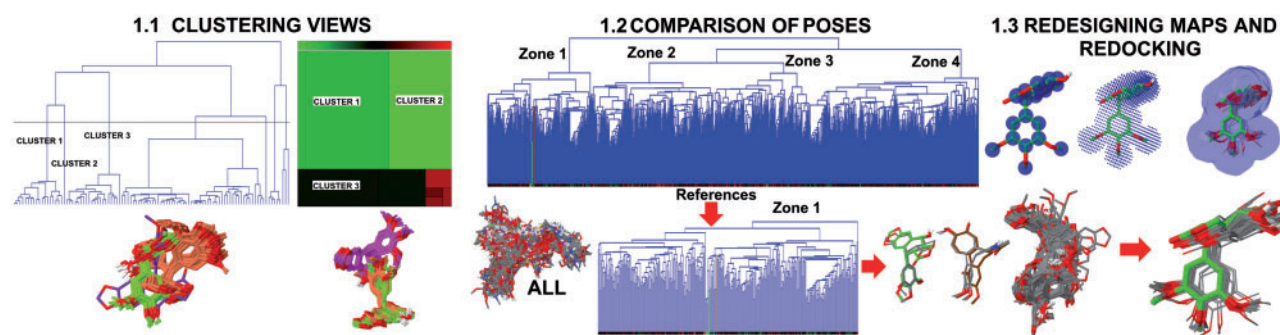


Fig. 1 Clustering views: dendrogram and box-like views (hierarchical and K-RMSD). Comparison of poses: Poses were not represented by their atoms but by their shortest distances to a fixed set of observers extracted from map files (see *Suppl. Material*). 1749 representatives, selected in the first step, were clustered, plus two references (green and brown lines). Redesigning maps and redocking: top: spheres are positioned based on a reference ligand (left) and the AutoDock A map values (center) were modified (right); bottom: initial docking results for the reference ligand (left), and the new results (right)

clustering algorithms are available within JADOPPT: the hierarchical (average, complete and single linkage) and the partitional algorithm (which was called K-RMSD, see *Suppl. Material*). The clustering results are presented as dendrograms, or box-like views for the K-RMSD (see *Fig. 1.1*). We have added a graphical interface to the molecular viewer Jmol (*Herráez, 2006*) for interactive 3D visualization of the poses selected on the dendrogram or box-like views. The interface allows the analysis and comparison of clusters through visual inspection of their representatives by clicking on any rectangle of the box-like visualization (*Fig. 1.1*), their binding energies, and the change of the automatically selected representatives. Moreover, the binding energies are color coded, (red-green for less to more affinity to the protein). On the other hand, the user can click on the branches in the dendrogram visualization for the analysis of clusters, as well as of individual molecules. Detailed information of the tools is provided in the manual.

The second task involves the conjoint clustering of multiple ligands from the poses selected in the previous step, with the aim of allowing a visual comparison of different molecules. The 3D comparison of poses of different molecules is a difficult task, as there is no direct correspondence between their atoms. Therefore, the atomic coordinates for the poses selected in step 1 were converted to a projection onto a fixed number of elements (called observers), so that all the molecules contained the same number of descriptors, which enabled the clustering calculation. We reasoned that the observers should come from the interaction grids, employed by AutoDock for energy evaluations. However, the number of grid points in the map files were too high for reasonably fast clustering computation. Therefore, we reduced them to map zones of sufficient volume to allocate atoms that contributed most to the binding energy (see *Suppl. Material*). Once the observers were selected, the scores were calculated based on the Molecular Similarity method (MS) proposed by Jain (*Jain, 2000*). Furthermore, the observers were selected to uniquely represent the force field atom types defined in AutoDock (*Morris et al., 2009*), as follows:

- Each atom type was considered independently.
- Atom types were grouped according to their similar interacting properties with the receptors, which allows conservative chemical atom replacements. Selected groups of AutoDock atom types were: 1) C, A, N; 2) OA, SA, NA; 3) F, Cl, Br, I; 4) HD; 5) e. (See *Suppl. Material*).
- All atom types were considered as a whole, which results in a mainly steric calculation of the similarity. Finally, an

agglomerative algorithm grouped the poses and generated an interactive dendrogram view (*Fig. 1.2*).

In the third task, we designed JADOPPT with the capability of interactively modifying the map files from AutoDock. This step was devised because groups of poses (*Fig. 1.2*) that have little likelihood of being chemically sound were found during visual analysis. Therefore, the map files could be modified to prevent the appearance of non-relevant branches in subsequent docking exercises, thus improving the sampling of more relevant zones of the chemical space (*Fig. 1.3* and *Suppl. Material*). Additionally, this map modification tool could also be applied to add pharmacophore-based structural requirements and/or information arising from structure-activity relationships, which can indirectly account for the flexibility of the target.

3 Results

For demonstration purposes, the tool has been applied to 17 600 poses resulting from the docking of colchicine and 10 analogues onto 16 tubulin models (see *Suppl. Material*, Case Study 1). *Fig. 1.1* shows the results of step 1: representative cluster options. The box-like view displays eight clusters found for the 100 poses of a single molecule (podophyllotoxin). The larger squares represent entropy favored clusters and are colored from green (enthalpy favored) to red in decreasing order of binding affinity. In the dendrogram, seven clusters are selected with the sliding grey bar, and their representative poses are shown in the structure viewer, and colored to make cluster comparison straightforward (clusters in green and orange, reference in purple). The dendrogram of *Fig. 1.2* reflects the clustering of the representative poses for the 11 ligands plus two reference compounds (green and brown). The dendrogram shows four large zones, with the first zone containing the two references. This suggests that the poses in this cluster are bound to tubulin in a similar way as the references from X-ray studies do. At the bottom, *Fig. 1.2* (ALL) shows 1751 poses and next to it, a zoom view of zone 1 displays where the references belong. Selecting the lines close to the references in the dendrogram shows the most similar representatives to the references, as shown on the right. *Fig. 1.3* shows the maps modification tool, designed to improve the docking results by modifying the interaction scores (see *Suppl. Material*). After redesigning two map files and re-docking the same structure that in *Fig. 1.1*, we can see that the results were improved (*Fig. 1.3* at the bottom) as some of the spurious poses are no longer present in the docking results.

4 Conclusion

We have presented an innovative visual tool for analyzing and comparing multiple docking results. Moreover, our approach reduced the analysis time and validation of docking trials. Several techniques and methods have been applied to target different tasks such as clustering, extraction and visualization of docking results, along with the comparison of different compounds through a hierarchical clustering display, such as the dendrogram, while providing enough flexibility to account for possibly diverse scenarios. In addition, a tool has been implemented with the aim of improving the docking results in subsequent docking campaigns by modifying the information provided to AutoDock. JADOPPT is a visual analytical tool that can complement the available tools for analyzing and refining the docking results. Ongoing work will expand the tool to other virtual screening platforms and further improve the processing capabilities.

Funding

Financial support was provided by the Consejería de Educación (Junta de Castilla y León) and FEDER funds (projects SA147U13 and SA030U16). C.G-P. thanks the Universidad Autónoma de Tamaulipas for a predoctoral fellowship.

Conflict of Interest: none declared.

References

- Jain,A.N. (2000) Morphological similarity: a 3D molecular similarity method correlated with protein-ligand recognition. *J. Comput. Aided Mol. Des.*, **14**, 199–213.
- Herráez,A. *et al.* (2006) Biomolecules in the computer: Jmol to the rescue. *Biochem. Mol. Biol. Educ.*, **34**, 255–261.
- Lill,M.A. and Danielson,M.L. (2011) Computer-aided drug design platform using PyMOL. *J. Comput. Aided Mol. Des.*, **25**, 13–19.
- Morris,G.M. *et al.* (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.*, **30**, 2785–2791.
- Morris,G.M. *et al.* (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.*, **19**, 1639–1662.
- Seeliger,D. and Groot,B. (2010) Ligand docking and binding site analysis with PyMOL and Autodock/Vina. *J. Comput. Aided Mol. Des.*, **24**, 417–422.
- Vaqué,M. *et al.* (2006) BDT: an easy-to-use front-end application for automation of massive docking tasks and complex docking strategies with AutoDock. *Bioinformatics*, **22**, 1803–1804.
- Zhang,S. *et al.* (2008) DOVIS: an implementation for high-throughput virtual screening using AutoDock. *BMC Bioinformatics*, **9**, 126.